

Stats Worksheet-1

- Q1) Answer= (a) True
- Q2) Answer= (a) Central Limit Theorem
- Q3) Answer= (b) Modeling bounded count data
- Q4) Answer= (d) All of the mentioned
- Q5) Answer= (c) Poisson
- Q6) Answer= (b) False
- Q7) Answer= (b) Hypothesis
- Q8) Answer= (a) 0
- Q9) Answer= (c) Outliers cannot conform to the regression relationship
- Q10) Answer= **Normal distribution**, also known as the **Gaussian distribution**, is a probability distribution that is **symmetric** about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean which also gives the shape of a bell and hence is also known as **bell shaped curve**.
In normal distribution **Mean, median and mode are all equal in value**
- Q11) Answer= There are **6** ways to handle missing data:
 - a) **Deleting rows** with missing values
 - b) Impute missing values for **continuous** variable
 - c) Impute missing values for **categorical** values

d) Using **algorithms** that support missing values

e) **Prediction** of missing values

f) Imputation using **deep learning** library

I'd like to recommend **two** data imputation techniques: **Average imputation** and **common point** imputation.

• Q12) Answer= **A/B testing** also know as **split testing** or **bucket testing** is a method of **comparing** two version of a **webpage** or **app** against each other to determine which one **performs better**.

• Q13) Answer= Mean imputation of missing data is **not acceptable** practice because mean variance **decreases** the **variance** of our data while **increasing bias**, as a result of **reduces variance** the model is **less accurate** and **confidence interval** is **narrower**.

• Q14) Answer= **Linear regression** analysis is used to **predict** the value of a **variable** based on the **value** of **another variable**. The variable we want to predict is called the **dependent variable** and the variable we are using to predict the other variable's value is called the **independent variable**.

Equation for linear regression is-

$Y = a + bx + e$, where ,

Y=dependent variable,

a= intercept

b= slope

x= independent variable

e= error

• Q15) Answer= There are **two** main branches of statistics.-

1) Descriptive statistics

2) Inferential statistics

